



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Using Large Diabetes Databases for Research

Citation for published version:

(on Behalf of the Scottish Diabetes Research Network Epidemiology Group), Wild, S, Fischbacher, C & McKnight, J 2016, 'Using Large Diabetes Databases for Research', *Journal of Diabetes Science and Technology*. <https://doi.org/10.1177/1932296816645120>

Digital Object Identifier (DOI):

[10.1177/1932296816645120](https://doi.org/10.1177/1932296816645120)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Diabetes Science and Technology

Publisher Rights Statement:

© 2016 Diabetes Technology Society

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Using Large Diabetes Databases for Research

Journal of Diabetes Science and Technology
2016, Vol. 10(5) 1073–1078
© 2016 Diabetes Technology Society
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1932296816645120
dst.sagepub.com



**Sarah Wild, FRCP(E)¹, Colin Fischbacher, FRCP(E)²,
and John McKnight, FRCP(E)³ (on Behalf of the Scottish
Diabetes Research Network Epidemiology Group)**

Abstract

There are an increasing number of clinical, administrative and trial databases that can be used for research. These are particularly valuable if there are opportunities for linkage to other databases. This paper describes examples of the use of large diabetes databases for research. It reviews the advantages and disadvantages of using large diabetes databases for research and suggests solutions for some challenges. Large, high-quality databases offer potential sources of information for research at relatively low cost. Fundamental issues for using databases for research are the completeness of capture of cases within the population and time period of interest and accuracy of the diagnosis of diabetes and outcomes of interest. The extent to which people included in the database are representative should be considered if the database is not population based and there is the intention to extrapolate findings to the wider diabetes population. Information on key variables such as date of diagnosis or duration of diabetes may not be available at all, may be inaccurate or may contain a large amount of missing data. Information on key confounding factors is rarely available for the nondiabetic or general population limiting comparisons with the population of people with diabetes. However comparisons that allow for differences in distribution of important demographic factors may be feasible using data for the whole population or a matched cohort study design. In summary, diabetes databases can be used to address important research questions. Understanding the strengths and limitations of this approach is crucial to interpret the findings appropriately.

Keywords

database, linkage, research, diabetes

Large diabetes databases with substantial population coverage can be generated from a variety of sources. These include directly from data collected as part of routine clinical care, data collected for quality registers, administrative data such as population estimates and death registrations, and research studies using both observational and trial designs. Each approach has advantages and disadvantages and understanding the limitations is essential for interpreting findings from research using different types of information. For example, although trial data are likely to represent the best quality data for assessing effectiveness of an intervention in terms of internal consistency, completeness and accuracy they are likely to be derived from a subgroup of the population that might therefore introduce bias and sample sizes are frequently too small to study rare outcomes. In contrast data collated from sources not designed primarily for research including clinical records, quality registers, and administrative data may be quicker and cheaper to obtain than trial data in settings where such systems are well established. These frequently include larger, more representative populations than trial-based databases and are at less risk of recall and observer bias that may develop from knowledge of research

questions among trial participants and staff. However these “secondary-use” databases are likely to include poorer quality data, frequently with missing or limited data on important potential confounding factors and findings from their analysis therefore need to be interpreted appropriately.

The focus of this review is on the use of databases derived from clinical records, quality registers and administrative data and covers some of the key issues in use and interpretation of these data. The recent publication of the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement¹ will encourage authors to be more explicit in describing their approach to using such data.

¹Usher institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

²Information Services Division, NHS National Services Scotland, Edinburgh, UK

³Metabolic Unit, Western General Hospital, Edinburgh, UK

Corresponding Author:

Sarah Wild, FRCP(E), Usher institute of Population Health Sciences and Informatics, University of Edinburgh, Medical School, Teviot Place, Edinburgh, Scotland, EH8 9AG, UK.
Email: Sarah.Wild@ed.ac.uk

Table 1. Advantages and Disadvantages of Using Large Databases for Research.

Advantages	Disadvantages
Relatively quick and cheap to use compared to primary data collection once set up	Effort and expense required to set up and maintain
Representative if population based	Data governance issues must be addressed
Further potential from linkage to other databases and for comparisons to other databases if definitions are harmonized	Completeness and accuracy must be described
Large numbers improve power for studying rare outcomes	May lack key variables
	Potential for bias
	Statistically significant differences may not be clinically significant

Outline of the Review

This review describes the Scottish approach, as described previously in detail,² to collecting data on almost the whole population of people with diagnosed diabetes in Scotland and for secondary use of the data for research and compares and contrasts the approaches used for other databases. In brief, the current Scottish Care Information (SCI)–Diabetes system has been developed over many years by clinical and technical experts to support clinical care, building on an agreed set of consistent definitions to define a core data set. SCI-diabetes has expanded over time to include call and recall for diabetic retinopathy screening and assessment of foot risk.

Relevant information is entered by clinical staff responsible for diabetes management in primary and secondary care as part of the process of care. They also gain access to information relevant to diabetes care entered by other health professionals. An annual summary of the data in the system is provided in the Scottish Diabetes Survey.³ Patients have access to their records through the My Diabetes My Way patient portal.⁴ Following approval from a national ethics committee and from the Public Benefit and Privacy Panel (a body that reviews the use of data for research for which consent has not been obtained from every individual), deidentified data are made available to researchers who have received approved data governance training. Key aspects of using the SCI-Diabetes database for research are considered below, with comparison to approaches used in other diabetes databases. Advantages and disadvantages of using large databases for research are summarized in Table 1.

Important Components of a Database

Diagnosis of Diabetes

Clearly the key entry criterion for a diabetes database should be a valid diagnosis of diabetes. This can however be challenging to establish. International bodies provide diagnostic criteria that change over time^{5,6} and are unlikely to be applied

uniformly in clinical practice. Few databases provide information on the criteria used to make a diagnosis. Several quality improvement registers such as the Swedish Diabetes Register and the databases that contribute to the Joint Asia Diabetes Evaluation (JADE) registry are predominantly based on registration of people with diagnosed diabetes by clinical staff.^{7,8} Electronic medical records of patients attending 1 or more centers such as those collated by Dr Mohan's Diabetes Specialities Centres in southern India provide a resource for clinical care and for research.⁹ The Cardiovascular Risk Reduction in South Asia Translation Trial is based on people with a diagnosis of diabetes on the basis of the 1999 World Health Organization criteria and data from participants have been used to produce several research papers beyond the original trial.¹⁰

The Scottish diabetes register derived from SCI-Diabetes is populated on the basis of diagnoses made in primary and secondary care using fasting (and, less commonly, postchallenge) glucose levels and random HbA1c values and the subsequent data entry and use of relevant disease codes in primary care databases and data entry into electronic clinical records after a hospital outpatient/ambulatory care visit. It is possible for SCI-Diabetes to feed relevant information collected in other settings into primary care databases. This is extremely helpful for audit and for providing information for the pay-for-performance scheme, the Quality Outcomes Framework, introduced into UK general practice in 2004. Daily linkage to the Community Health Index master patient index provides frequent updating of dates of registration and deregistration with Scottish general practice as well as dates of death. The SCI-diabetes database is validated in clinical practice using responses to annual invitations to diabetic retinopathy screening. For example, in 2010 191 571 individuals were invited for screening of whom 2% were suspended from the screening program and diabetes register because the person was subsequently found not to have diabetes.¹¹

Extracts from the clinical database that have been deidentified for use in research have been validated against hospital admissions mentioning diabetes¹² or against sulphonylurea prescriptions (unpublished data) and this work suggests that over 99% of people with diagnosed diabetes in Scotland are included in both clinical and research databases. As a consequence misclassification bias of diagnosed diabetes is not likely to be a major problem. As type 2 diabetes may be asymptomatic there are likely to be people with undiagnosed diabetes who are not included in the SCI-diabetes database. The most recent estimates, derived from estimates of prevalence of both diagnosed and undiagnosed diabetes from models using screening data compared to estimates of prevalence of diagnosed diabetes, suggest that approximately 15% of diabetes may be undiagnosed in Scotland.¹³

Other diabetes databases use different approaches to validation of diagnosis of diabetes. For example, the case definition used by QRESEARCH in a database derived from 240 UK primary care practices and 4 million patients is based on

the presence of a specific disease code or more than 2 prescriptions for drugs used to treat diabetes (insulin or oral agents).¹⁴ The Danish Diabetes Register is also based on a combination of prescribing data and disease codes. Validation suggests that completeness based on these criteria is $\geq 95\%$, although date of diagnosis of diabetes may be unreliable and there may be overestimation of the number of people with diabetes from glucose measurement alone.¹⁵

Some studies use coding of diabetes in hospital records to identify people with diabetes (for example)¹⁶ and the completeness of this approach can be expected to vary depending on choice of codes, outcome of interest and hospital.¹² When using International Classification of Disease codes it is important to be aware of changes in coding rules between different revisions. For example there were important changes between the 9th and 10th revisions in Rule 3, which allows a condition originally identified as contributing to death to be assigned as the underlying cause of death if it is a direct consequence of the condition originally identified as the underlying cause.¹⁷ For example, under this rule when the underlying cause was recorded as nephrotic syndrome it would be reassigned to diabetes if diabetes was mentioned anywhere else in the certificate. In Scotland and England this change took place in 2000 and the number of deaths attributed to diabetes increased by approximately 4% as a consequence.¹⁸

In New Zealand and Canada algorithms have been developed to combine data from several sources, building on the capture-recapture approach that has been used previously to create population based diabetes registers.¹⁹⁻²² A systematic review and meta-analysis of 6 studies which reported sensitivity and specificity of 2 physician claims or 1 hospital discharge abstract record within a 2-year period as a case definition of diabetes concluded that the definition missed up to 20% of cases of diabetes and falsely identified 2% of people as having diabetes.²³

Population based prescribing or dispensing databases can be used to identify people with drug treated diabetes as the majority of drugs used to treat diabetes, with the exception of metformin, are not used for other indications. For example, Norwegian prescribing data have been used to validate a pediatric diabetes register and to estimate the prevalence of diabetes,²⁴ although these data do not cover people in nursing homes, hospitals or other institutions so are likely to underestimate prevalence. Identifying the first date of prescription can be used to identify incident diagnoses although this may not be an accurate proxy for the date of diagnosis of diabetes. Prescribing data can be challenging to use as consequence of different coding systems for drug classes, unclear information about start and stop dates, the potential for patients not to fill prescriptions, and the fact that a reasonably large proportion of people with type 2 diabetes in many populations are not prescribed drug treatment. In Scotland in 2009, for example, we estimated that approximately 18 % of the 196 000 people over 39 years of age with a diagnosis of type

2 diabetes did not fill a prescription for treatment of diabetes, either because it was not prescribed or because they did not collect the prescription (unpublished data). This proportion is likely to vary between populations.

Type and Duration of Diabetes

Many diabetes databases do not distinguish between type 1 and type 2 diabetes and those that do may rely on inconsistent approaches applied to defining type of diabetes used by different clinical staff, resulting in misclassification. The legacy of the previous nomenclature of insulin-dependent diabetes and non-insulin-dependent diabetes means that the most frequent misclassification is of people who receive insulin treatment for type 2 diabetes receiving a label of type 1 diabetes. It is possible to develop algorithms using a variety of different sources of information including age at diagnosis of diabetes and treatment patterns to classify or validate type of diabetes.²⁵ The QRESEARCH group defines type 1 diabetes as age ≤ 35 years at diagnosis and evidence of treatment with insulin, including devices. As date of diagnosis of diabetes may not be recorded reliably, the algorithms may use other information from prescribing, biochemical or retinopathy screening records to validate date of diagnosis of diabetes.²⁶ Duration of diabetes is an important variable for many analyses and the absence of reliable information on date of diagnosis from many databases limits their value for research.

Other Factors

The availability of valid data on potential confounding factors, effect modifiers and outcomes of interest also influences the value of databases for audit and research. Although assays for key parameters such as HbA1c, creatinine, and estimated glomerular filtration rate are expected to be standardized, there is still the potential for important differences over time and between laboratories and extensive investigation and data manipulation may be required to ensure that meaningful comparisons can be made. Each variable that is used is likely to require cleaning to remove implausible data. Further challenges arise from irregular recording of key variables and large amounts of missing data. Modern computing capacity mean that methods for handling missing data in very large databases are now feasible and the validity of using complete case analysis (that is a restricted sample size of people with complete data) can now be assessed.

Opportunities and Challenges Relating to Use of Large Databases for Research

The availability of diabetes databases for research has increased in recent years, partly following recognition of the value of registries for managing individual care and for improving equity of outcomes of care across populations.

The amount of time and effort required to set up such systems should not be underestimated and partly explains the limited number of population-based registers, particularly in less developed countries. It is also important to understand the value of secondary uses of the data for research purposes if local requirements for data governance can be met (as discussed further below). There will be further opportunities for using these data if comparability across countries can be improved. Harmonization of data set definitions allows cross-country comparisons and pooling of data to examine rare outcomes. However, harmonization requires extensive work; collaborative projects such as EUBIROD are supporting the development of new diabetes data sets so that colleagues can benefit from the years of work and experience in setting up such systems in Scotland and other countries.²⁷

Once the challenges of ensuring that the data in diabetes databases are internally valid have been met, as described above, it becomes possible to consider the wider application of the data for research, either from linkage to other databases, where available, or to make comparisons with the general population. For example, diabetes databases can be used to identify diabetes as an outcome as an efficient way of conducting longer term follow-up of participants in trials²⁸ or observational studies such as birth cohorts.²⁹

There may be data governance and logistical issues to address before anonymous linkage of data is possible without explicit individual consent from all people whose information is included in a database. There are marked differences between countries in the perceived balance between the advantages of having population level data for use by health services and researchers and the need to protect confidentiality. Proposed new European legislation might mean that disease registers based on implied consent are no longer allowed in the future so that data only become available from selected populations. Clearly protecting patient confidentiality and informing people how information about them might be used is essential and when setting up a new disease register informed consent should be obtained from individuals where feasible.

In Scotland a linked database of acute hospital admissions, cancer registrations, deaths and admissions to psychiatric hospitals is available³⁰ and the use of a unique national patient identifier (the Community Health Index number) on records, including those in the diabetes database, allows linkage prior to deidentification for research purposes. Examples of how the linked database have been used are available from various studies of morbidity and mortality.^{26,31-34} It is also possible to link to other disease registers in Scotland, such as the renal register.³⁵ Several of these studies have made comparisons with either the general population or the whole population without diabetes although this approach is usually limited to the ability to adjust for confounding by age, sex and an area-based measure of socioeconomic status. Swedish colleagues undertaking similar studies

have made comparisons between people with diabetes and controls randomly selected from the general population who were matched by age, sex, and geographical area.³⁶ The potential for ascertainment bias in studies of the association between diabetes and cancer, in which the diagnoses of both conditions occur within a similar time frame, either because of opportunistic testing or increased surveillance should be noted.³⁷

Pharmaco-epidemiological studies are also possible with use of prescribing data and often require international collaboration to provide sufficient data, particularly for new drugs, requiring development of methods to share aggregated data and avoid sharing of individual-level data across countries.^{38,39} Observational studies of drug exposure are prone to time-related bias. Many studies describing apparent benefits of metformin on cancer incidence have demonstrated such bias resulting in exaggerated estimates of the benefits of metformin compared to studies that have avoided such bias.^{40,41} A further potential for misleading findings arises from the potential for confounding by indication or channeling bias in which risk of the outcome influences choice of drug. Approaches to tackling this issue include use of propensity scores but it is important to be aware of the limitations of these approaches.⁴²

Conclusions

Several well-established and validated diabetes databases exist that provide the basis for valuable and efficient primary and secondary uses when data governance and methodological challenges have been addressed. Although setting up such databases takes a considerable amount of resources and time there are opportunities to learn from the experience of people responsible for existing databases.

Abbreviations

JADE, Joint Asia Diabetes Evaluation; RECORD, REporting of studies Conducted using Observational Routinely-collected health Data; SCI, Scottish Care Information.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Benchimol EI, Smeeth L, Guttman A, et al. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLOS MED*. 2015;12:e1001885.

2. Cunningham S, McAlpine R, Leese G, et al. Using web technology to support population-based diabetes care. *J Diabetes Sci Technol*. 2011;5:523-534.
3. Diabetes in Scotland. Scottish Diabetes Survey 2014. Available at: <http://www.diabetesinscotland.org.uk/Publications.aspx?catId=3>. 2016. Accessed January 25, 2016.
4. NHS Scotland. My diabetes my way. 2016. Available at: <http://www.mydiabetesmyway.scot.nhs.uk/>. Accessed January 25, 2016.
5. American Diabetes Association. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care*. 1997;20:1183-1197.
6. World Health Organization, International Diabetes Federation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. Geneva: World Health Organization; 2006.
7. Gudbjornsdottir S, Cederholm J, Nilsson PM, Eliasson B. The National Diabetes Register in Sweden: an implementation of the St. Vincent Declaration for Quality Improvement in Diabetes Care. *Diabetes Care*. 2003;26:1270-1276.
8. Chan JC, Ozaki R, Luk A, et al. Delivery of integrated diabetes care using logistics and information technology—the Joint Asia Diabetes Evaluation (JADE) program. *Diabetes Res Clin Pract*. 2014;106(suppl 2):S295-S304.
9. Pradeepa R, Prabu AV, Jebarani S, Subhashini S, Mohan V. Use of a large diabetes electronic medical record system in India: clinical and research applications. *J Diabetes Sci Technol*. 2011;5:543-552.
10. Shah S, Singh K, Ali MK, et al. Improving diabetes care: multi-component cardiovascular disease risk reduction strategies for people with diabetes in South Asia—the CARRS multi-center translation trial. *Diabetes Res Clin Pract*. 2012;98:285-294.
11. NHS Scotland. SCI-Diabetes. Available at: <http://www.sci-diabetes.scot.nhs.uk/project-overview/>. 2016. Accessed January 25, 2016.
12. Anwar H, Fischbacher CM, Leese GP, Lindsay RS, McKnight J, Wild SH. Assessment of the under-reporting of diabetes in hospital admission data: a study from the Scottish Diabetes Research Network Epidemiology Group. *Diabet Med*. 2011;28:1514-1519.
13. ScotPHO. Undiagnosed diabetes. 2016. Available at: <http://www.scotpho.org.uk/health-wellbeing-and-disease/diabetes/data/undiagnosed-diabetes>. Accessed January 25, 2016.
14. Hippisley-Cox J, Ryan R. Diabetes in the United Kingdom: Analysis of QRESEARCH data. Available at: http://www.qresearch.org/Public_Documents/DataValidation/Diabetes%20in%20the%20UK%20analysis%20of%20QRESEARCH%20data.pdf. 2016. Accessed January 25, 2016.
15. Green A, Sortso C, Jensen PB, Emneus M. Validation of the Danish National Diabetes Register. *Clin Epidemiol*. 2015;7:5-15.
16. Vamos EP, Millett C, Parsons C, Aylin P, Majeed A, Bottle A. Nationwide study on trends in hospital admissions for major cardiovascular events and procedures among people with and without diabetes in England, 2004-2009. *Diabetes Care*. 2012;35:265-272.
17. World Health Organization. ICD-10: international statistical classification of diseases and related health problems: tenth revision—2nd ed. 2016. Available at: http://www.who.int/classifications/icd/ICD-10_2nd_ed_volume2.pdf. Accessed February 1, 2016.
18. Office for National Statistics. International classifications of diseases, 10th revision—ICD10 for mortality. 2016. Available at: <http://www.ons.gov.uk/ons/guide-method/classifications/international-standard-classifications/icd-10-for-mortality/index.html>. Accessed January 25, 2016.
19. Jo EC, Drury PL. Development of a virtual diabetes register using information technology in New Zealand. *Healthc Inform Res*. 2015;21:49-55.
20. Harvey JN, Craney L, Kelly D. Estimation of the prevalence of diagnosed diabetes from primary care and secondary care source data: comparison of record linkage with capture-recapture analysis. *J Epidemiol Community Health*. 2002;56:18-23.
21. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care*. 2002;25:512-516.
22. Chen G, Khan N, Walker R, Quan H. Validating ICD coding algorithms for diabetes mellitus from administrative data. *Diabetes Res Clin Pract*. 2010;89:189-195.
23. Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. *PLOS ONE*. 2013;8:e75256.
24. Berg C, Strom H. The Norwegian Prescription Database (NorPD) as a data source for diabetes research. *Norsk Epidemiologi*. 2013;23(1):109-110.
25. de LS, Khunti K, Belsey J, et al. A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. *Diabet Med*. 2010;27:203-209.
26. Livingstone SJ, Looker HC, Hothersall EJ, et al. Risk of cardiovascular disease and total mortality in adults with type 1 diabetes: Scottish registry linkage study. *PLOS MED*. 2012;9:e1001321.
27. Cunningham SG, Carinci F, Brillante M, et al. Core standards of the EUBIROD project. Defining a European diabetes data dictionary for clinical audit and healthcare delivery. *Methods Inf Med*. 2016;55:166-176.
28. Bhopal RS, Douglas A, Wallia S, et al. Effect of a lifestyle intervention on weight change in south Asian individuals in the UK at high risk of type 2 diabetes: a family-cluster randomised controlled trial. *Lancet Diabetes Endocrinol*. 2014;2:218-227.
29. Robertson L, Harrild K. Maternal and neonatal risk factors for childhood type 1 diabetes: a matched case-control study. *BMC Public Health*. 2010;10:281.
30. ISD Scotland. ACaDMe. 2016. 25-1-2016.
31. Hothersall EJ, Livingstone SJ, Looker HC, et al. Contemporary risk of hip fracture in type 1 and type 2 diabetes: a national registry study from Scotland. *J Bone Miner Res*. 2014;29:1054-1060.
32. Walker JJ, Livingstone SJ, Colhoun HM, et al. Effect of socioeconomic status on mortality among people with type 2 diabetes: a study from the Scottish Diabetes Research Network Epidemiology Group. *Diabetes Care*. 2011;34:1127-1132.
33. Walker JJ, Brewster DH, Colhoun HM, et al. Type 2 diabetes, socioeconomic status and risk of cancer in Scotland 2001-2007. *Diabetologia*. 2013;56:1712-1715.

34. Jackson CA, Jones NR, Walker JJ, et al. Area-based socioeconomic status, type 2 diabetes and cardiovascular mortality in Scotland. *Diabetologia*. 2012;55:2938-2945.
35. Bell S, Fletcher EH, Brady I, et al. End-stage renal disease and survival in people with diabetes: a national database linkage study. *QJM*. 2015;108:127-134.
36. Lind M, Svensson AM, Kosiborod M, et al. Glycemic control and excess mortality in type 1 diabetes. *N Engl J Med*. 2014;371:1972-1982.
37. Johnson JA, Carstensen B, Witte D, Bowker SL, Lipscombe L, Renehan AG. Diabetes and cancer (1): evaluating the temporal relationship between type 2 diabetes and cancer incidence. *Diabetologia*. 2012;55:1607-1618.
38. Colhoun HM, Livingstone SJ, Looker HC, et al. Hospitalised hip fracture risk with rosiglitazone and pioglitazone use compared with other glucose-lowering drugs. *Diabetologia*. 2012;55:2929-2937.
39. Levin D, Bell S, Sund R, et al. Pioglitazone and bladder cancer risk: a multipopulation pooled, cumulative exposure analysis. *Diabetologia*. 2015;58:493-504.
40. Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care*. 2012;35:2665-2673.
41. Gandini S, Puntoni M, Heckman-Stoddard B, et al. Metformin and cancer risk and mortality: a systematic review and meta-analysis taking into account biases and confounders. *Cancer Prev Res (Phila)*. 2014;7:867-885.
42. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. *BMJ*. 2013;347:f6409.